

# Asymptotic and numerical methods for metastable events in stochastic gene networks

Jay Newby\*

December 30, 2014

## Abstract

A general class of stochastic gene expression models with self regulation is considered. One or more genes randomly switch between regulatory states, each having a different mRNA transcription rate. The gene or genes are self regulating when the proteins they produce affect the rate of switching between regulatory states. Under weak noise conditions, the deterministic forces are much stronger than fluctuations from gene switching and protein synthesis. Metastable transitions, such as bistable switching, can occur under weak noise conditions, causing dramatic shifts in the expression of a gene. A general tool used to describe metastability is the quasi stationary analysis (QSA). A large deviation principle is derived so that the QSA can explicitly account for random gene switching without using an adiabatic limit or diffusion approximation, which are unreliable and inaccurate for metastable events. This allows the existing asymptotic and numerical methods that have been developed for continuous Markov processes to be used to analyze the full model.

## 1 Introduction

Stochasticity in gene regulation circuits is known to cause rare extreme shifts in the expression of a gene, which can have a profound effect on the behavior of a cell [7, 19]. This leads to the question of how cells coordinate and regulate different sources of biochemical fluctuations, or noise, to function within a genetic circuit. Rare, noise-induced dynamical shifts in a stochastic process are known as *metastable events*. For example, Brownian motion in a double well potential, where the fluctuations are weak compared to the force of the potential, displays bistable switching. In general, metastable events occur when fluctuations are weak compared to deterministic forces, and the stochastic process is said to be under *weak noise* conditions. The challenge for stochastic modeling is to predict and explain the possible metastable behaviors and offer a testable hypothesis by quantifying the timescale on which those events are likely to occur.

Standard methods for computing transition rates in chemical reactions cannot be applied to a gene expression model when it includes reactions that involve one or more genes. For chemical systems with a large number  $N$  of molecules, the stochastic model converges to the mass action model in the large  $N$  limit. Transition rate approximations for chemical systems generally apply to weak noise conditions where  $N$  is large but finite. A deterministic mass action description is not valid for a gene expression model because there can be as few as one copy of a given gene. Hence, we cannot assume weak noise conditions by assuming that  $N$  is large. Even if we can find a way to define weak noise for the gene expression model, the standard transition rate approximation does not apply.

A deterministic description can be obtained using a separation of time scales if the gene reactions are fast compared to other reactions; this is known as an

---

\*Mathematical Bioscience Institute, Ohio State University, 1735 Neil Ave. Columbus, OH 43210

adiabatic limit. For example, a gene that rapidly switches between on and off states would, in the adiabatic limit, have an effective transcription rate scaled by the fraction of time spent in the on state. A stochastic gene regulation model can then be said to be under weak noise conditions when gene switching is fast but not infinitely fast.

Thus, there is a separation of timescales, where gene switching is the fast reaction and protein synthesis/degradation is the slow reaction. Almost all of the approaches to date resolve the gene switching problem by either taking the adiabatic limit [2] or using a diffusion approximation [2, 32, 8, 17], but diffusion approximations of metastable phenomena, in general, are known to be highly unreliable, even for relatively simple birth-death processes [5]. The adiabatic diffusion approximation (also known as stochastic averaging or stochastic quasi-steady state) can be particularly unreliable [25, 14, 23]. Other approaches are either not generally applicable [1] or result in a difficult to solve high dimensional problem [28, 33]. We propose an accurate approximation that exploits the separation of timescales without ignoring gene switching or using diffusion approximations of any kind.

The study of metastable behavior in probability theory is known as *large deviation theory* [10, 9], and has many mathematical similarities to the path integral description of quantum field theory. Large deviation theory is mathematically rigorous and establishes an important link between the path distribution and the stationary probability density function, but for most chemical systems and continuous Markov processes, the WKB approximation is the simplest and most direct means of calculating certain weak noise approximations.

The Quasistationary analysis (QSA) uses the WKB method to approximate the stationary probability density function, which also yields information about the long time behavior of the process [29]. One useful byproduct of the WKB approximation is the so-called *most probable path* (MPP). Under weak noise conditions, the probability that the process takes a particular path from point A to point B is sharply peaked along a MPP. The likelihood of deviating from the MPP is an exponentially decreasing function of the magnitude of the deviation. In other words, stochastic trajectories are highly likely to closely follow the MPP. A MPP is a statistic, similar to the mode, of the probability distribution of stochastic trajectories, (i.e., a probability distribution in a function space). Although they describe a stochastic process, MPPs themselves are not stochastic. In general, there are two kinds of MPPs. If a deterministic trajectory connects point A to point B, then the MPP is the deterministic trajectory. If the transition is noise induced, such as a metastable transition, MPPs build *action* and become more improbable as the path gets longer. MPPs can tell us how different metastable transitions occur; they can be thought of as describing noise induced dynamical behavior. By combining the stationary density and MPP approximations, the mean first exit time for a metastable events (such as bistable switching) can also be approximated using a spectral projection method. To summarize, the QSA is based on the WKB approximation and can be used to obtain asymptotic approximations of: (i) the stationary probability density function, (ii) MPPs, and (iii) mean first exit times for metastable transitions. How to apply the QSA to processes that have both large- $N$ -type species and fast (adiabatic) species with low copy number is an unsolved problem.

Wolynes and coworkers showed how the quantum mechanical path integral formulation of spin boson systems could be applied to study metastability in stochastic gene expression models that include gene switching [28, 33]. They obtained a path integral description in the full space of variables (both slow and fast) [33], but the result is a high dimensional problem that can be difficult to solve. In this paper, we present an approach that exploits the separation of time scales. Recently, the QSA has been adapted to get approximations of the stationary density and mean switching times [24, 23, 22], but the link to a path integral formulation is ambiguous because the WKB approximation does not uniquely

determine a Hamiltonian. (The leading order problem from the WKB expansion is a static Hamilton-Jacobi equation.) Two important numerical algorithms (the geometric minimum action method [12, 31] and the ordered upwind method [4, 22]) critically depend on the correct formulation of the Hamiltonian.

Recent advances in this area have been made using probability theory. Using a *large deviation principle*, one can derive the action functional, which is the basis for approximations of metastable behavior. Freidlin and Wentzell [10] present a rigorous proof of a large deviation principle for the case where the fast process does not depend on the slow variables (i.e., no gene regulation), and Kifer [16] extended the result to the fully coupled case without noise in the slow variables (the slow process by itself is deterministic). The model is called *piecewise deterministic* if the fast process is a discrete Markov process and the slow process is deterministic. A large deviation principle was developed for piecewise deterministic processes using path integrals [3]. The piecewise deterministic case is relevant for a gene expression model in the large  $N$  limit where the protein concentration changes deterministically in between random gene state transitions [24]. In this paper, we consider the case where the slow process also has noise from protein synthesis and degradation. Specifically, we assume that the synthesis rate is much larger than the degradation rate so that the protein is under large- $N$ -type weak noise conditions and can be treated as a continuous concentration.

One of the weaknesses of the large deviation theory approach is that the resulting approximations are logarithmically asymptotic. For example, if  $P_\epsilon(x)$  is a probability density function that depends on a small parameter  $\epsilon \ll 1$ , a large deviation principle yields an approximation of the form,  $\epsilon \log(P_\epsilon(x)) \sim -\Phi(x) + O(\epsilon)$ , which we write as  $P_\epsilon(x) \stackrel{\text{ln}}{\sim} e^{-\Phi(x)/\epsilon}$ . The QSA yields an asymptotic approximation that includes a pre exponential factor  $K(x)$  so that  $P_\epsilon(x) \sim K(x)e^{-\Phi(x)/\epsilon}$ .

In this paper, we show how to resolve the ambiguity in the WKB approximation, which allows the QSA to be applied to models of stochastic gene networks that include multiple types of weak noise (large- $N$ -type weak noise and adiabatic-type weak noise). For simplicity we ignore the dependence on mRNA copy number and focus on gene state switching and protein synthesis and degradation. The author has developed the QSA for single gene models that include mRNA using transform methods [21], and we anticipate that the theory can be expanded to include gene networks in future work. We develop a path description that takes advantage of the separation timescales without using an adiabatic reduction (stochastic averaging) or any other form of diffusion approximation. This can be thought of as an averaging-like procedure that yields highly accurate approximations of many relevant quantities, including the distribution of the fast gene state. We then combine the results of large deviation theory and the QSA. That is, for practical application, we show how to augment the WKB method to obtain all relevant quantities without the need to consider path integrals.

The paper is organized as follows. First, we describe a general stochastic model of gene expression in Section 2. Then, in Section 3 we present the WKB approximation and explain how to resolve the ambiguity in defining the Hamilton-Jacobi equation. In Section 4 we show how a single mathematical concept (the asymptotic approximation of an integral using Laplace's method) can be used to derive a large deviation principle, which establishes how to correctly apply the WKB approximation. Finally, in section 5 we illustrate the results with the simple example of the *toggle switch* model (also known as the *mutual repressors* model) [15, 24].

## 2 Model

Let  $s = 1, \dots, N_s$  label the state of one or more genes. For simplicity, we assume that protein synthesis is directly coupled to the gene state and ignore mRNA (for more on applying the QSA to a model that does consider mRNA see [1, 21].)

Assume that the gene or genes are responsible for synthesizing  $d > 0$  protein species  $X_1, \dots, X_d$ , and let  $\mathbf{x} \in \mathcal{D} \subset \mathbb{R}^d$  be the vector of protein concentrations. Protein species  $X_k$  is synthesized and degraded according to the reaction,



where  $\sigma_k(s)/\epsilon$  is the gene state dependent production rate and  $\gamma_k$  is the degradation rate. Assume that  $0 < \epsilon \ll 1$  is a small parameter so that  $\mathbf{x}$  can be treated as a continuous variable. We assume that there are no direct interactions between the different protein species, but extending the analysis to account for such reactions is straightforward.

Given a fixed gene state  $s$ , the Master equation describing the evolution of the probability density function  $P(\mathbf{x}, t)$  is

$$\begin{aligned} \frac{\partial}{\partial t} P(\mathbf{x}, t) = L_{\mathbf{x}}^{\epsilon} P \equiv & \frac{1}{\epsilon} \sum_{k=1}^d \sum_{x'_k \in \mathcal{D}} \left[ \sigma_k(s) (\delta_{x'_k, x_k - \epsilon} - \delta_{x'_k, x_k}) \right. \\ & \left. + \gamma_k (\delta_{x'_k, x_k + \epsilon} - \delta_{x'_k, x_k}) x'_k \right] P(\mathbf{x}', t) \end{aligned} \quad (2)$$

Assume that the  $X_1, \dots, X_d$  regulate the activity of the gene or genes so that the transition rates are functions of  $\mathbf{x}$ . Given a fixed concentration  $\mathbf{x}$ , the gene state is described by the Master equation

$$\frac{\partial}{\partial t} P(s, t) = \frac{1}{\epsilon} A^{\mathbf{x}} P \equiv \frac{1}{\epsilon} \sum_{s'=1}^{N_s} A_{ss'}^{\mathbf{x}} P(s', t), \quad (3)$$

where  $A^{\mathbf{x}}$  is the transition rate matrix, assumed to be irreducible so that there is a steady-state distribution  $\rho(s|\mathbf{x})$  that satisfies  $A^{\mathbf{x}} \rho = 0$ .

For the coupled process, where both  $s$  and  $\mathbf{x}$  are random, the evolution of the joint probability density function  $P(s, \mathbf{x}, t)$  is governed by the Master equation,

$$\frac{\partial}{\partial t} P(s, \mathbf{x}, t | s_0, \mathbf{x}_0, t_0) = \mathcal{L}_{\epsilon} P \equiv \frac{1}{\epsilon} A^{\mathbf{x}} P + L_{\mathbf{x}}^{\epsilon} P. \quad (4)$$

In the limit  $\epsilon \rightarrow 0$  we obtain the deterministic system,

$$\frac{dx_k}{dt} = f_k(\mathbf{x}) \equiv \sum_{s=0}^{N_s} \rho(s|\mathbf{x}) \sigma_k(s) - \gamma_k x_k. \quad (5)$$

Fixed points are written as  $\mathbf{x}_{\text{fp}}$ , defined by  $f_k(\mathbf{x}_{\text{fp}}) = 0$ , for all  $k = 1, \dots, d$ .

### 3 WKB analysis

First we consider a special case, one that is well known in the chemical literature [6, 27]. Suppose the gene state is fixed constant. The WKB approximation of the stationary probability density function is obtained using the ansatz  $P(\mathbf{x}) \stackrel{\text{ln}}{\simeq} e^{-\Phi(\mathbf{x})/\epsilon}$ . Substituting the ansatz into (46) and collecting leading order terms in  $\epsilon$  yields the PDE  $H(s, \mathbf{x}, \mathbf{p}) = 0$ , where  $\mathbf{p} \equiv \nabla_{\mathbf{x}} \Phi$  and

$$H(s, \mathbf{x}, \mathbf{p}) \equiv \sum_{k=1}^d \left[ \sigma_k(s) (e^{p_k} - 1) + \gamma_k x_k (e^{-p_k} - 1) \right]. \quad (6)$$

For the coupled process, where  $s$  is also random, the situation is more complicated. A popular approach (which is known to cause significant errors [23, 33]) is to take the adiabatic limit and use the averaged Hamiltonian  $\bar{H}(\mathbf{x}, \mathbf{p}) = H(\sum_s s \rho(s|\mathbf{x}), \mathbf{x}, \mathbf{p})$ .

A much more accurate approximation can be obtained as follows. The WKB approximation of the stationary probability density function is obtained using the ansatz

$$P(s, \mathbf{x}) = \mathcal{N} K(\mathbf{x}) r(s|\mathbf{x}) e^{-\Phi(\mathbf{x})/\epsilon}, \quad (7)$$

where  $\mathcal{N}$  is a normalization factor,  $r(s|\mathbf{x})$  is the conditional distribution of the gene state given  $\mathbf{x}$ , and  $\Phi(\mathbf{x})$  is called the *quasipotential*. The pre exponential factor  $K(\mathbf{x})$  can be interpreted as a normalization factor for  $r$  and is determined at higher order. For completeness, we provide the details on how to compute  $K(\mathbf{x})$  in Appendix B.

Substituting (7) into  $\mathcal{L}_{\mathbf{x}}^\epsilon P = 0$  and collecting leading order terms in  $\epsilon$  yields

$$\sum_{s'=0}^{N_s} [A_{ss'}^\mathbf{x} + \delta_{ss'} H(s, \mathbf{x}, \Phi'(\mathbf{x}))] r(s'|\mathbf{x}) = 0. \quad (8)$$

This can be rewritten in matrix form as

$$Qr = [A^\mathbf{x} + \text{diag}\{H(s, \mathbf{x}, \nabla_\mathbf{x} \Phi(\mathbf{x}))\}]r = 0. \quad (9)$$

It is not immediately clear what the Hamiltonian is since the above expression depends on  $r$ , which is also unknown. One idea is to define the Hamiltonian to be  $\mathcal{H}(\mathbf{x}, \mathbf{p}) = \det(Q)$ . But this definition is not the only one that we can make, for example, we could also define the Hamiltonian to be an eigenvalue of  $Q$ . Then  $r$  is simply the eigenvector (which we hope is positive since it approximates a probability distribution) corresponding to a zero eigenvalue. Both of these definitions should yield the same WKB approximation if we can solve the static Hamilton-Jacobi equation,  $\mathcal{H}(\mathbf{x}, \nabla_\mathbf{x} \Phi) = 0$ . The solution to the Hamilton-Jacobi equation can be written in terms of the method of characteristics [26]. Setting  $\mathbf{p} = \nabla_\mathbf{x} \Phi$ , the Hamilton-Jacobi equation can be rewritten as the system of ODEs,

$$\dot{\mathbf{x}} = \nabla_{\mathbf{p}} \mathcal{H}(\mathbf{x}, \mathbf{p}), \quad \dot{\mathbf{p}} = -\nabla_{\mathbf{x}} \mathcal{H}(\mathbf{x}, \mathbf{p}). \quad (10)$$

Now we ask a simple question. Are MPPs also solutions to (10)? The answer is certainly not evident from the WKB expansion.

We are then faced with a new question. If we have multiple definitions of  $\mathcal{H}$  that yield the same WKB approximation, do they also yield the same characteristics, and if the characteristics are different, which set correspond to MPPs? Clearly, the solution surface  $\nabla_\mathbf{x} \Phi(\mathbf{x}(t, \theta))$  can have different parameterizations even with the same initial data (i.e., the curves of constant  $\theta$  need not be the same for different choices of  $\mathcal{H}$ ). This ambiguity is resolved in Section 4 using a large deviation principle, and the correct choice is to define  $\mathcal{H}$  as the principal eigenvalue of  $Q$ ; it is real and simple, it is larger than the real part of the remaining eigenvalues, and its eigenvector  $r$  is strictly positive. The *Perron–Frobenius Theorem* guarantees that the principal eigenvalue exists with  $r > 0$  for all  $p$  [22]. Moreover, it follows from the WKB approximation (7) that the conditional probability distribution for the gene state is just the appropriately normalized right eigenvector (evaluated at  $\mathbf{p} = \nabla \Phi(\mathbf{x})$ ).

The correct choice of  $\mathcal{H}$  also has significant consequences for certain numerical algorithms for computing MPPs and the quasipotential  $\Phi$  that are based on the large deviation principle [12, 31, 4, 22]. Another shortcoming of making the wrong choice for  $\mathcal{H}$  is that these algorithms depend on  $\mathcal{H}(\mathbf{x}, \mathbf{p})$  being a convex function of  $\mathbf{p}$ . Choosing  $\mathcal{H}$  according to the large deviation principle guarantees that it is convex. One way to compute MPPs and  $\Phi$  is to use a numerical ODE integrator to solve (10), and this method does not require  $\mathcal{H}$  be convex [24, 17]. However, there are two significant drawbacks of this method. First, there is little control over which points  $\mathbf{x}$  the solution is evaluated at; often one is interested in approximating  $\Phi$  on a discrete grid. Each MPP is generated with a different initial condition and a shooting method is required to find the MPP with a specific end point. Second, the shooting method rapidly becomes

computationally expensive as the dimension  $d$  increases. A better choice for computing MPPs is the *geometric minimum action method* (GMAM), which is an iterative scheme to compute the MPP between a given start and end point [12, 31]. The quasipotential  $\Phi(\mathbf{x})$  can be computed on a discrete grid with the *ordered upwind method* (OUM), which is a finite difference scheme for solving the Hamilton-Jacobi equation [4, 22]. The GMAM and OUM requires  $\mathcal{H}(\mathbf{x}, \mathbf{p})$  and its various first and second order partial derivatives.

Because it is an eigenvalue of a matrix, it is not possible in general to explicitly calculate  $\mathcal{H}$  except in special cases. However, there is an efficient iterative scheme, called the Collatz method, to numerically compute the principal eigenvalue, the details are included in Appendix C. If we can numerically compute  $\mathcal{H}$  then we need only know the coefficients in the characteristic polynomial of  $Q$  (which can be explicitly calculated in general) to evaluate the various partial derivatives of  $\mathcal{H}$ . Write the characteristic equation for  $Q$  as  $\Omega(\lambda) \equiv a_{N_s} \lambda^{N_s} + \dots + a_1 \lambda + a_0 = 0$ , where the coefficients  $a_j$  are functions of  $\mathbf{x}, \mathbf{p}$ . Differentiating the characteristic equation and setting  $\lambda = \mathcal{H}$  yields

$$\nabla \mathcal{H}(\mathbf{x}, \mathbf{p}) = - \frac{\sum_{n=0}^{N_s} \mathcal{H}^n \nabla a_n(\mathbf{x}, \mathbf{p})}{\sum_{n=1}^{N_s} n \mathcal{H}^{n-1} a_n(\mathbf{x}, \mathbf{p})}, \quad (11)$$

Higher derivatives can be similarly defined. We have found in practice that simply setting  $\mathcal{H} = 0$  in the above formula seems to yield well behaved numerical solutions, but justifying this choice in terms of stability requires more investigation.

## 4 Deriving the large deviation principle

Large deviation estimates are based on the idea of formulating Laplace's method specifically for probability theory. Recall that Laplace's method is used to approximate certain integrals. For example, suppose we have a positive function  $F(x)$  for which there is a single minimum at  $x_m$  such that  $F'(x_m) = 0$  with  $\lim_{x \rightarrow \pm\infty} F(x) = \infty$ . Laplace's method yields the approximation

$$\int_{-\infty}^{\infty} e^{-F(x)/\epsilon} dx \stackrel{\ln}{\simeq} e^{-F(x_m)/\epsilon}.$$

The idea is that the integral takes its largest contribution at the point  $x_m$  where  $F(x)$  is at its minimum because the integrand is exponentially decreasing away from  $x_m$ .

To see how this works in the context of probability theory, consider the random variable  $X_\epsilon$ . We want to approximate the distribution with

$$\Pr[X_\epsilon \in (x, x + dx)] \stackrel{\ln}{\simeq} dx e^{-L(x)/\epsilon}, \quad (12)$$

for some function  $L(x)$ . Let  $p \in \mathbb{R}$ . Using (12) we notice that

$$\left\langle \exp \left\{ \frac{pX_\epsilon}{\epsilon} \right\} \right\rangle \stackrel{\ln}{\simeq} \int_{-\infty}^{\infty} \exp \left\{ \frac{px - L(x)}{\epsilon} \right\} dx.$$

Using Laplace's method, we know that the above integral obtains its largest contribution at  $h(p) = \sup_{x \in \mathbb{R}} \{xp - L(x)\}$  so that  $\langle \exp\{pX_\epsilon/\epsilon\} \rangle \stackrel{\ln}{\simeq} \exp\{h(p)/\epsilon\}$ . Hence, we can define the function,

$$h(p) \equiv \lim_{\epsilon \rightarrow 0} \epsilon \log \left\langle \exp \left\{ \frac{pX_\epsilon}{\epsilon} \right\} \right\rangle. \quad (13)$$

Assume that  $h$  is continuous and convex for all  $p \in \mathbb{R}$  with  $h(0) = 0$ . Then, we can define the function  $L$  and its relationship to  $h$  through the *Legendre transform*



$\mathcal{T}$ ,

$$L(x) \equiv \mathcal{T}[h(p)] = \sup_{p \in \mathbb{R}} \{xp - h(p)\} \quad (14)$$

$$h(p) = \mathcal{T}[L(x)] = \sup_{x \in \mathbb{R}} \{xp - L(x)\}. \quad (15)$$

If  $h(p)$  is a convex function of  $p$  then  $L$  is a convex function of  $x$ , and their derivatives are monotonic functions. After differentiating each of the above expressions, we observe that the maximizers for (14) and (15) are given implicitly by  $x = h'(p)$ ,  $p = L'(x)$ , respectively. Hence, the procedure for Laplace's method is to first compute  $h$  using (13). If the limit exists then the probability density function can be approximated by (12), where  $L(x)$  is given by (14). This procedure can be generalized to apply Laplace's method to path integrals.

For simplicity, assume that there is a single slow species so that  $d = 1$  and  $0 \leq x < \infty$ . The generalization for multiple slow species is presented in Section 4.2. Discretize time with  $t_j = t_0 + j\Delta t$ ,  $j = 0, \dots, N + 1$ . Let  $\{s_j, x_j\}$  be a discretized path where  $s_j = s(t_j)$  and  $x_j = x(t_j)$ . Assume that the end points ( $j = 0$  and  $j = N + 1$ ) of the path are fixed with  $s_{N+1} = s$  and  $x_{N+1} = x$ .

The probability density function satisfying the Master equation can be written in terms of a path integral, which is a slightly different version of the standard integral form of the Chapman–Kolmogorov equation [11]. If the path consists of a single interior point (i.e.,  $N = 1$ ) then we have the Chapman–Kolmogorov equation,

$$P(s, x, t | s_0, x_0, t_0) = \sum_{s_1=0}^{N_s} \int_0^\infty dx_1 P(s, x, t | s_1, x_1, t_1) P(s_1, x_1, t_1 | s_0, x_0, t_0). \quad (16)$$

Compounding over multiple interior points along a path ( $N > 1$ ), we have

$$P(s, x, t | s_0, x_0, t_0) = \sum_{s_1, \dots, s_N} \int_0^\infty dx_1 \cdots \int_0^\infty dx_N \mathcal{P}[\{s_j\}, \{x_j\}], \quad (17)$$

where  $\mathcal{P}$  is the joint distribution over the path. Using the Markov property, the path distribution can be written as the product

$$\mathcal{P}[\{s_j\}, \{x_j\}, \{t_j\}] = \prod_{j=1}^N P(s_j, x_j, t_j | s_{j-1}, x_{j-1}, t_{j-1}). \quad (18)$$

We are interested in “averaging out” (not to be confused with stochastic averaging or adiabatic reduction) the dependence on  $s$ , hence we seek

$$\mathcal{P}[\{x_j\}] = \sum_{s_0, \dots, s_{N+1}} \mathcal{P}[\{s_j\}, \{x_j\}].$$

Formally we take the continuum limit  $N \rightarrow \infty$  with  $\Delta t \rightarrow 0$ , and we assume that the random gene state  $s(T)$  evolves on the fast timescale  $T = t/\epsilon$ . We further assume that  $x(t)$  and  $p(t)$  are slowly varying functions on the fast timescale (i.e.,  $x'(t)$  and  $p'(t)$  are  $O(1)$ ). The marginal path distribution  $\mathcal{P}[x(t)]$  can be approximated as follows. The generalization of (13) for the path distribution is

$$\mathcal{H}(x, p) = \lim_{\epsilon \rightarrow 0} \epsilon \log \left\langle \exp \left\{ \int_0^{t/\epsilon} H(s(\tau), x, p) d\tau \right\} \right\rangle, \quad (19)$$

where

$$H(s, x, p) = \sigma(s)(e^p - 1) + \gamma x(e^{-p} - 1).$$

Note that (19) is mathematically rigorous for the case where there is no noise in the slow process (i.e., the protein synthesis and degradation is deterministic) and

for the adiabatic limit (i.e., no noise from gene switching). As formally justified in Appendix A, (19) is a straightforward generalization for the case where both noise sources are present.

It is worth taking a moment to compare (19) to the adiabatic result. The adiabatic limit yields a very different result, namely  $\bar{H}(x, p) = \lim_{\epsilon \rightarrow 0} H(\langle s(t/\epsilon) \rangle, x, p)$ . In general, using the adiabatic limit or a diffusion approximation yields a Hamiltonian that agrees with (19) only for  $|p| \ll 1$ .

The generalization of (14) is

$$L[x, \dot{x}] = \sup_{p \in \mathbb{R}} [\dot{x}p - \mathcal{H}(x, p)], \quad \mathcal{H}(x, p) = \sup_{\dot{x} \in \mathbb{R}} [\dot{x}p - L[x, \dot{x}]], \quad (20)$$

where  $L$  is called the *Lagrangian* and the variable  $p$  is called the *conjugate momentum*. The maximizers are defined implicitly by

$$\dot{x} = \frac{\partial}{\partial p} \mathcal{H}(x, p), \quad p = \frac{\partial}{\partial \dot{x}} L[x, \dot{x}]. \quad (21)$$

Using (19), the generalization of (12) for the marginal path distribution is the large deviation principle,

$$\mathcal{P}[x(t)] \stackrel{\text{ln}}{\simeq} \exp \left\{ -\frac{1}{\epsilon} \mathcal{S}[x(t)] \right\}, \quad (22)$$

where  $\mathcal{S}[x(t)] = \int_{t_0}^t L[x(\tau), \dot{x}(\tau)] d\tau$  is called the *action functional*; it is a measure of how likely a given stochastic trajectory is. It follows from the definition of  $L$  and the property  $\mathcal{H}(x, 0) = 0$  that  $\frac{d}{dt} \mathcal{S}[x(t)] \geq 0$ . As the action increases, the likelihood of observing a stochastic trajectory decreases.

#### 4.1 Most probable paths and WKB

Combining (16) with (22) we can approximate the marginal probability density function  $p(x, t)$  with the path integral over the set of continuous functions in  $\mathbf{C}_{x_0}^x = \{x(t) \in \mathbf{C}_1 : x(0) = x_0, x(t) = x\}$ ,

$$P(x, t | x_0, t_0) \stackrel{\text{ln}}{\simeq} \int_{\mathbf{C}_{x_0}^x} \mathcal{D}[x(t)] \exp \left\{ -\frac{1}{\epsilon} \mathcal{S}[x(t)] \right\}. \quad (23)$$

Rather than try to evaluate an integral on a function space, let us use Laplace's method to approximate the integral. The largest contribution to the above path integral comes from the MPP, defined as the path that minimizes the action functional,

$$x_{\text{MP}}(t) \equiv \arg \inf_{x(t) \in \mathbf{C}_{x_0}^x} \int_{t_0}^t L[x(\tau), \dot{x}(\tau)] d\tau. \quad (24)$$

Stochastic trajectories are highly likely to follow a MPP. The probability of a  $O(1)$  deviation from a MPP is exponentially small. Using Laplace's method again, we have the approximation

$$P(x, t | x_0, t_0) \stackrel{\text{ln}}{\simeq} \exp \left\{ -\frac{1}{\epsilon} \mathcal{S}[x_{\text{MP}}(t)] \right\}. \quad (25)$$

Using the calculus of variations, one can show that the MPP defined by (24) is a solution to Hamilton's equations (10), derived using the WKB method! That is, an MPP is a solution to the Euler-Lagrange equation,

$$\frac{d}{dt} \frac{\partial}{\partial \dot{x}} L[x, \dot{x}] = \frac{\partial}{\partial x} L[x, \dot{x}]. \quad (26)$$

The above can be equivalently written as

$$\frac{d}{dt} \left( \dot{x} \frac{\partial}{\partial \dot{x}} L[x, \dot{x}] - L[x, \dot{x}] \right) = 0. \quad (27)$$



Using (20), we have the first integral,

$$\dot{x}p - L[x, \dot{x}] = \mathcal{H}(x, p) = \text{Const.} \quad (28)$$

Hence,  $\mathcal{H}$  is constant along a MPP. The first equation in (10), namely  $\dot{x} = \frac{\partial \mathcal{H}}{\partial p}$ , simply follows from (21). The second equation is derived as follows. Using the second equation in (21) and (26), we have

$$\dot{p} = \frac{d}{dt} \frac{\partial}{\partial \dot{x}} L[x, \dot{x}] = \frac{\partial}{\partial x} L[x, \dot{x}]. \quad (29)$$

We therefore need to show that  $\frac{\partial}{\partial \dot{x}} L[x, \dot{x}] = -\frac{\partial}{\partial x} \mathcal{H}(x, p)$  on MPPs. Using (20) and writing the maximizer as  $p(x)$  (i.e., the implicit solution of  $\frac{\partial}{\partial x} \mathcal{H}(x, p) = \dot{x}$ ) as a function of  $x$  yields

$$\frac{\partial}{\partial \dot{x}} L[x, \dot{x}] = \frac{\partial p}{\partial x} \dot{x} - \frac{\partial p}{\partial x} \frac{\partial}{\partial p} \mathcal{H}(x, p(x)) - \frac{\partial}{\partial x} \mathcal{H}(x, p(x)). \quad (30)$$

Then, it follows from the first equation in (21) that  $\frac{\partial}{\partial \dot{x}} L[x, \dot{x}] = -\frac{\partial}{\partial x} \mathcal{H}(x, p(x))$ .

The final connection to the WKB solution comes by setting  $\mathcal{H} = 0$ . As noted above  $\mathcal{H} = \text{Const}$  along MPPs. If we restrict trajectories to the set of MPPs for which  $\mathcal{H} = 0$ , the action (22) becomes

$$\mathcal{S}[x_{\text{MP}}(t)] = \int_{t_0}^t p(\tau) \dot{x}_{\text{MP}}(\tau) d\tau. \quad (31)$$

The connection between the large deviation principle and the WKB approximation (7) of the stationary probability density function is as follows. A MPP that starts at the stable fixed point  $x_A$  requires  $t_0 \rightarrow -\infty$  as  $x_0 \rightarrow x_A$ , because  $x_A$  is an unstable fixed point of (10) (even though it is a stable fixed point of the deterministic system). In fact, the deterministic system is recovered from (10) by setting  $p = 0$ , which means that deterministic flows exist on the stable manifold of the higher dimensional system. Assume that the process is time autonomous so that the stationary density can be written as  $P(x) = \lim_{t_0 \rightarrow -\infty} P(x, t|x_A, t_0)$ . We can approximate the stationary density by taking the limit  $t_0 \rightarrow -\infty$  and  $x_0 \rightarrow x_A$  in (25) and (31) (assuming  $x(t)$  and  $p(t)$  satisfy (10)) to get

$$P(x) \propto \exp \left\{ -\frac{1}{\epsilon} \int_{-\infty}^t p(\tau) \dot{x}(\tau) d\tau \right\} \equiv e^{-\Phi(x)/\epsilon}, \quad (32)$$

which is consistent with the WKB approximation (7). Notice that the conjugate momentum  $p$  can also be interpreted as  $p = \frac{d\Phi}{dx}$ . It follows from (31) and (32) that the action determines the quasipotential with  $\Phi(x(t)) = \mathcal{S}[x_{\text{MP}}(t)]$ . Recall that  $\mathcal{S}[x(t)]$  is a non decreasing function of time. If  $\dot{x}_{\text{MP}}(t) \neq 0$  for some fixed  $t$ , it is easy to show that  $\frac{d}{dt} \mathcal{S}[x_{\text{MP}}(t)] = 0$  (and consequently  $p = 0$ ) if and only if  $\dot{x} = f(x)$ . In other words, the action and the quasipotential are flat when  $x_{\text{MP}}(t)$  is tangent to a deterministic trajectory. This is always true at fixed points  $x_{\text{fp}}$ .

## 4.2 Multiple slow species

For multiple slow species  $\mathbf{x} \in \mathbb{R}^d$ ,  $d > 1$ , the Hamiltonian  $\mathcal{H}(\mathbf{x}, \mathbf{p})$  is a function of the vector  $\mathbf{p} = (p_1, \dots, p_d)$ . The Hamiltonian is defined as the principal eigenvalue of the matrix  $Q_{s,s'} = A_{s,s'}^{\mathbf{x}} + \delta_{s,s'} H(s, \mathbf{x}, \mathbf{p})$  and  $r(s|\mathbf{x}) > 0$  is its eigenvector. The Lagrangian (20) becomes

$$L[\mathbf{x}, \dot{\mathbf{x}}] = \sup_{\mathbf{p} \in \mathbb{R}^d} \{ \mathbf{p} \cdot \dot{\mathbf{x}} - \mathcal{H}(\mathbf{x}, \mathbf{p}) \}. \quad (33)$$

MPPs satisfy Hamilton's equations (10).

## 5 Example: Toggle Switch

We turn now to a simple example to illustrate the results. Consider two genes,  $G_X$  and  $G_Y$ , responsible for producing proteins  $X$  and  $Y$ , which have concentrations  $x$  and  $y$ . We assume that both genes can be turned on or one of the two can be turned off while the other is on. Let  $s = 1$  correspond to  $G_X = \text{on}$ ,  $G_Y = \text{off}$ ; let  $s = 2$  correspond to  $G_X = G_Y = \text{on}$ ; and let  $s = 3$  correspond to  $G_X = \text{off}$ ,  $G_Y = \text{on}$ . Assume that a repressor formed by  $X$  turns  $G_Y$  off and a repressor formed by  $Y$  turns  $G_X$  off. Once off, each gene turns back on at a rate  $b$  when regulatory sequences unbind from the repressor. The transition rate matrix for the gene state is,

$$A^{(x,y)} = \begin{bmatrix} -b & x^\alpha & 0 \\ b & -x^\alpha - y^\alpha & b \\ 0 & y^\alpha & -b \end{bmatrix}, \quad (34)$$

where  $\alpha$  is a parameter that controls the amount of nonlinearity in the dependence of the repressor concentration on protein concentration. (For simplicity, we assume that the repressor concentrations are an instantaneous function of  $x$  and  $y$ .) The scaled protein synthesis rates are  $\sigma_X(\{1, 2, 3\}) = \{0, \sigma, \sigma\}$  and  $\sigma_Y(\{1, 2, 3\}) = \{\sigma, \sigma, 0\}$ . We set the characteristic time scale to the average lifetime of a protein so that the degradation rates are  $\gamma_X = \gamma_Y = 1$ . For  $\alpha \geq 2$  there is a range of values of  $b$  for which the deterministic limit (5) is bistable, having two stable fixed points  $(x_A, y_A)$  and  $(y_A, x_A)$ , separated by an unstable saddle located on the center line  $y = x$ . MPPs generated numerically using the GMAM are shown in Fig. 1 along with level curves of  $\Phi(x, y)$  computed using the OUM. The  $x, y$  symmetry is evident from the level curves and MPPs. Sur-

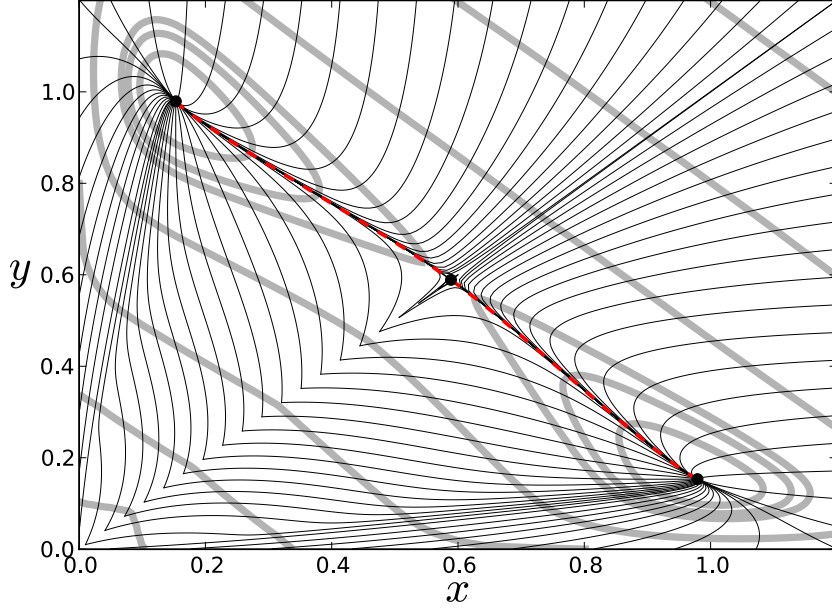


Figure 1: Several MPPs that start from each stable fixed point are shown (thin lines) for  $\alpha = 2$ ,  $b = 0.15$ , and  $\sigma = 1$ . Red dashed lines show the MPEPs connecting the stable fixed points to the saddle. Level curves of the quasipotential  $\Phi(x, y)$  are shown as thick lines.

rounding the two stable fixed points are closed level curves corresponding to a potential well. The dashed curves show the MPP from the stable fixed points to the saddle located on the center line. This curve is often referred to as the *most probable exit path* (MPEP) because it is the path most likely taken to transition from one of the stable fixed points to the other. There are many other MPPs that reach the center line, but along the center line, the quasipotential is minimized

at the saddle, which means that stochastic trajectories are most likely to cross the center line at the saddle.

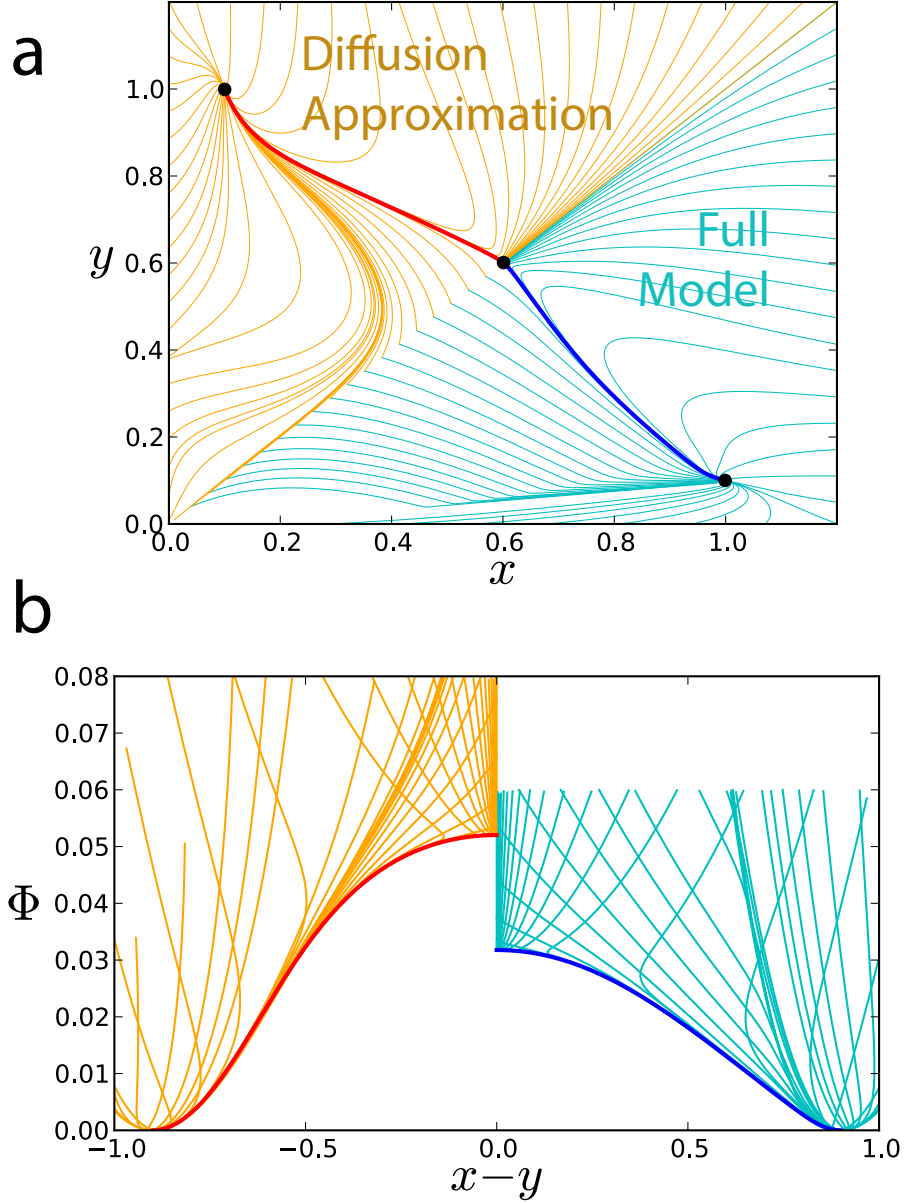


Figure 2: (a) MPPs computed using the GMAM for the full model ( $y < x$ ) and for the diffusion approximation ( $y > x$ ). The red and blue curves show the MPEP for the diffusion approximation and the full model, respectively. (b) The quasipotential along each of the MPPs shown in (a). Parameter values are  $\alpha = 3$ ,  $b = 0.15$ , and  $\sigma = 1$ .

It is interesting to see what differences a diffusion approximation has with these results. A complete diffusion approximation of the toggle switch model, that includes noise effects from gene state switching and protein synthesis/degradation, was developed in [15]. In Fig. 2(a) we show the MPPs derived from the full model in the lower half of the domain  $y < x$  and the analogous MPPs obtained from the diffusion approximation on the upper  $y > x$  portion of the domain. We expect the diffusion approximation to be accurate in a small neighborhood of the stable fixed points. After comparing the two approximations, we see that there are significant differences between the two sets of MPPs for small values of  $x$  and  $y$ . The MPEPs (red and blue curves in Fig. 2) are also slightly different.

To understand the main source of error in the diffusion approximation, we examine the difference in the quasipotential. Along a MPP, the quasipotential is given by

$$\Phi(\mathbf{x}(t)) = \int_0^t \mathbf{p}(\tau) \cdot \dot{\mathbf{x}}(\tau) d\tau, \quad (35)$$

where  $\mathbf{x}(t)$  and  $\mathbf{p}(t)$  satisfy (10). In Fig. 2(b) we show the quasi potential computed along each of the MPPs shown in Fig. 2(a). Notice a significant difference in the quasipotential at the saddle point, which has an exponentially large effect on the Kramers rate approximation of the mean exit time (not shown).

## 6 Discussion

We have derived a systematic, efficient, and practical extension of the QSA, which can be used to obtain reliable approximations of the stationary probability density function, MPPs, mean switching times, and the distribution of exit points. We emphasize that our approach does not make the mistake of using an adiabatic limit or a diffusion approximation, which are unreliable and inaccurate for metastable events. We have shown how the same asymptotic and numerical methods that exist to analyze a diffusion approximation can be applied to the full model. We therefore conclude that *there is little to no benefit to be gained by using a diffusion approximation*.

The matrix  $Q$  that determines  $\mathcal{H}$  can be obtained using the simple and direct WKB approximation, and in situations where the principal eigenvalue of  $Q$  cannot be explicitly calculated, an efficient numerical scheme can be used. The correct formulation of  $\mathcal{H}$  guarantees that it is a convex function of  $\mathbf{p}$ , a property that is required for numerical schemes such as the OUM and GMAM.

## A Large deviation principal

For a model with only adiabatic-type weak noise, having no noise in the slow variable, (19) becomes

$$\mathcal{H}(x, p) = \lim_{\epsilon \rightarrow 0} \epsilon \log \left\langle \exp \left\{ \int_0^{t/\epsilon} p f(S, x) d\tau \right\} \right\rangle, \quad (36)$$

where  $\frac{dx}{dt} = f(s, x)$  is the deterministic motion for a given fixed state  $s$ . The validity of the above is rigorously established in [16] for the case where  $S$  is a continuous Markov process, a Markov chain, or a combination of both. For comparison, one can derive from the gene expression model a limiting process without noise in the slow variable. By carefully rescaling the protein synthesis and degradation rates, one can derive a limiting process in which the protein concentration changes deterministically in between jumps in the gene state [24, 23]. In this limit (36) is recovered from (19). That is, we have that  $H(s, x, p) \rightarrow pf(s, x)$ , which is consistent with the Hamiltonian of a deterministic process.

We state as a conjecture that (19) is a straightforward generalization of (36) for the case where there is noise in the slow variable. In this section, we present a formal justification for this conjecture, using arguments that follow closely with the ideas found in the proof of (36).

We define the slow process as

$$X(t) = X(0) + \epsilon \left[ Y^+ \left( \frac{1}{\epsilon} \int_0^t \sigma(S(\epsilon^{-1}\tau)) d\tau \right) - Y^- \left( \frac{\gamma}{\epsilon} \int_0^t X(\tau) d\tau \right) \right],$$

where  $Y^\pm(t)$  are independent unit Poisson processes. The fast process is  $S(\epsilon^{-1}t)$ , with transition rate matrix  $A^x$ . Discretize the  $S, x(t)$  dependent slow process with

$X_j = X(t_j)$ , with increments  $\Delta X_j = X_j - X_{j-1}$ . We seek an approximation of the marginal path distribution  $\mathcal{P}[x(t)]$  having the form

$$\mathcal{P}[X(t) \approx x(t)] \stackrel{\text{ln}}{\simeq} \exp \left\{ -\frac{1}{\epsilon} \int_0^t L[x(\tau), \dot{x}(\tau)] d\tau \right\}. \quad (37)$$

Let  $X'_j = \Delta X_j / \Delta t$ . Then using the discrete version of (37) we have that

$$\left\langle \exp \left\{ \frac{1}{\epsilon} \sum_{j=1}^N p_j X'_j \Delta t \right\} \right\rangle \stackrel{\text{ln}}{\simeq} \int_{-\infty}^{\infty} dx'_1 \cdots \int_{-\infty}^{\infty} dx'_N \exp \left\{ \frac{1}{\epsilon} \sum_{j=1}^N (p_j x'_j - L[x_j, x'_j]) \Delta t \right\}.$$

The largest contribution to each integral comes from

$$\mathcal{H}(x_j, p_j) = \sup_{x'_j \in \mathbb{R}} \{p_j x'_j - L[x_j, x'_j]\},$$

Hence, the Hamiltonian should satisfy

$$\int_{t_0}^t \mathcal{H}(x(\tau), p(\tau)) d\tau = \lim_{\epsilon \rightarrow 0} \epsilon \log \left\langle \exp \left\{ \frac{1}{\epsilon} \int_{t_0}^t p(\tau) X'(\tau) d\tau \right\} \right\rangle. \quad (38)$$

A more useful expression than (38) is derived as follows

First, we evaluate the mean over jumps in  $X$  conditioned on  $S$ . To leading order in  $\Delta t$ , the propagator density function for  $\Delta X$  (the probability density of jumping from  $x$  to  $x + y$  in the time interval  $(t, t + \Delta t)$ ) is

$$T(y|\Delta t, s, x) \sim \delta(y) + \frac{\Delta t}{\epsilon} [\sigma(s)(\delta(y - \epsilon) - \delta(y)) + \gamma x(\delta(y - \epsilon) - \delta(y))].$$

Taking the mean of each  $\exp\{\epsilon^{-1} p_j \Delta X_j\}$  over all  $j = 0, \dots, N$  yields

$$\begin{aligned} & \int_{-\infty}^{\infty} dy_1 \cdots \int_{-\infty}^{\infty} dy_N \prod_{j=1}^N T(y_j|\Delta t, s, x_j) e^{p_j y_j / \epsilon} \\ &= \prod_{j=1}^N \left( 1 + \frac{\Delta t}{\epsilon} H(s, x_j, p_j) \right) \sim \exp \left\{ \frac{1}{\epsilon} \sum_{j=1}^N H(s, x_j, p_j) \Delta t \right\} + o(\Delta t), \end{aligned} \quad (39)$$

where

$$H(s, x, p) = \sigma(s)(e^p - 1) + \gamma x(e^{-p} - 1).$$

To complete the evaluation of (38), we must average over the fast process  $S(\epsilon^{-1}t)$ . Formally we take the continuum limit  $N \rightarrow \infty$  with  $\Delta t \rightarrow 0$ , and we assume that  $x(t)$  and  $p(t)$  are slowly varying functions on the fast timescale  $T = t/\epsilon$ . In the continuum limit we have

$$\lim_{\substack{N \rightarrow \infty \\ \Delta t \rightarrow 0}} \left\langle \exp \left\{ \frac{1}{\epsilon} \sum_{j=1}^N p_j \frac{\Delta X_j}{\Delta t} \Delta t \right\} \right\rangle \stackrel{\text{ln}}{\simeq} \left\langle \exp \left\{ \frac{1}{\epsilon} \int_0^t H(S(\epsilon^{-1}\tau), x(\tau), p(\tau)) d\tau \right\} \right\rangle. \quad (40)$$

Combining (38) and (40) yields,

$$\int_0^t \mathcal{H}(x(\tau), p(\tau)) d\tau = \lim_{\epsilon \rightarrow 0} \epsilon \log \left\langle \exp \left\{ \frac{1}{\epsilon} \int_0^t H(S(\epsilon^{-1}\tau), x(\tau), p(\tau)) d\tau \right\} \right\rangle. \quad (41)$$

Notice that (41) differs drastically from the adiabatic result, which is essentially

$$\begin{aligned} \int_0^t \bar{H}(x(\tau), p(\tau)) d\tau &= \lim_{\epsilon \rightarrow 0} \epsilon \log \exp \left\{ \frac{1}{\epsilon} \int_0^t H(\langle S(\epsilon^{-1}\tau) \rangle, x(\tau), p(\tau)) d\tau \right\} \\ &= \int_0^t H(\lim_{T \rightarrow \infty} \langle S(T) \rangle, x(\tau), p(\tau)) d\tau. \end{aligned}$$

Then, differentiating both sides with respect to time yields

$$\bar{H}(x(\tau), p(\tau)) = H(\lim_{T \rightarrow \infty} \langle S(T) \rangle, x(\tau), p(\tau)),$$

where

$$\lim_{T \rightarrow \infty} \langle S(T) \rangle = \sum_s s \rho(s|x), \quad \sum_{s'} A_{s,s'}^x \rho(s'|x) = 0, \quad \sum_s \rho(s|x) = 1.$$

Finally, we note that the formula (19) given in the previous section is much simpler than (41). A formal derivation of this formula using a multiple timescales argument is as follows. Change variables to the fast timescale  $T = t/\epsilon$  to get

$$\epsilon \int_0^T \mathcal{H}(x(t), p(t)) dT' \sim \epsilon \log \left\langle \exp \left\{ \int_0^T H(S(T'), x(t), p(t)) dT' \right\} \right\rangle,$$

For fixed  $t$  we have that  $T \rightarrow \infty$  as  $\epsilon \rightarrow 0$ . Divide through by  $\epsilon T$ , set  $x(t) = x$  and  $p(t) = p$ , and take the limit  $\epsilon \rightarrow 0$  with  $t$  fixed to get

$$\mathcal{H}(x, p) = \lim_{T \rightarrow \infty} \frac{1}{T} \log \left\langle \exp \left\{ \int_0^T H(S(T'), x, p) dT' \right\} \right\rangle.$$

The above mean assumes that  $x$  is fixed constant so that

$$\mathcal{H}(x, p) = \lim_{T \rightarrow \infty} \frac{1}{t} \log \left( \sum_{s'} U_{s,s'}(t) \right), \quad (42)$$

where  $U(t) = \exp\{tQ^T\}$  and  $Q = A^x + \text{diag}\{H(s, x, p)\}$ .

We now show that (42) converges to the principal eigenvalue of the matrix  $Q$ . Although a general proof is only a few lines (see [10]), we offer a more descriptive argument, valid for the case where  $Q$  is diagonalizable. Suppose that the matrix  $Q$  has eigenvalues  $\lambda_j$ , right eigenvectors  $r_j$ , and left eigenvectors  $l_j$ . Let the principal eigenvalue be  $\lambda_0$  with  $r_0 > 0$ ,  $l_0 > 0$ , and  $\lambda_0 > \Re(\lambda_j)$ ,  $j = 1, \dots, N_s$ . Define the constants  $c_j = \mathbf{1}^T r_j$ . Notice that  $c_0 > 0$  since  $r_0 > 0$ . We have

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{1}{t} \log(\exp\{tQ^T\} \mathbf{1}) &= \lim_{t \rightarrow \infty} \frac{1}{t} \log \sum_{j=0}^{N_s} l_j e^{\lambda_j t} c_j \\ &= \lim_{t \rightarrow \infty} \frac{1}{t} \left[ \log(\mathbf{1} e^{\lambda_0 t}) + \log \left( l_0 c_0 + \sum_{j=1}^N l_j e^{-(\lambda_0 - \lambda_j)t} c_j \right) \right] \\ &= \lambda_0 \mathbf{1} + \lim_{t \rightarrow \infty} \frac{1}{t} \log(l_0 c_0) = \lambda_0 \mathbf{1}. \end{aligned}$$

## B The pre exponential factor

From the leading order WKB solution we have  $\Phi(\mathbf{x})$  and  $r(s|\mathbf{x})$ , which satisfies  $\sum_{s'} Q_{ss'} r(s'|\mathbf{x}) = 0$ . Define the left null vector  $l(s|\mathbf{x})$  satisfying  $\sum_{s'} Q_{s's} l(s'|\mathbf{x}) = 0$ , and assume that  $l$  is normalized so that  $\sum_s l(s|\mathbf{x}) r(s|\mathbf{x}) = 1$ . Collecting second order terms in the WKB expansion and applying a solvability condition yields the prefactor equation,

$$\begin{aligned} &\sum_{k=1}^d \frac{\partial K}{\partial x_k} \sum_{s=1}^{N_s} l(s; \mathbf{x}) \frac{\partial J}{\partial p_k}(s, \mathbf{x}, \nabla \Phi) \\ &= K \sum_{s=1}^{N_s} l(s|\mathbf{x}) \sum_{k=1}^d \left[ \frac{\partial^2 J}{\partial p_k \partial x_k}(s, \mathbf{x}, \nabla \Phi) + \frac{1}{2} \sum_j^d \frac{\partial^2 \Phi}{\partial x_j \partial x_k} \frac{\partial^2 J}{\partial p_j \partial p_k}(s, \mathbf{x}, \nabla \Phi) \right] \end{aligned} \quad (43)$$

where

$$J(s, \mathbf{x}, \mathbf{p}) \equiv H(s, \mathbf{x}, \mathbf{p})r(s|\mathbf{x}).$$

Using the equation for the characteristics (10) we have that

$$\frac{dx_k}{dt} = \frac{\partial}{\partial p_k} \mathcal{H}(\mathbf{x}, \mathbf{p}) = \sum_{s=1}^{N_s} l(s|\mathbf{x}) \frac{\partial J}{\partial p_k}(s, \mathbf{x}, \nabla \Phi). \quad (44)$$

Using the chain rule, we have

$$\frac{d}{dt} K(\mathbf{x}(t)) = \sum_{k=1}^d \frac{\partial K}{\partial x_k} \frac{dx_k}{dt},$$

and it follows that along characteristics,

$$\frac{dK}{dt} = K \sum_{s=1}^{N_s} l(s|\mathbf{x}) \sum_{k=1}^d \left[ \frac{\partial^2 J}{\partial p_k \partial x_k}(s, \mathbf{x}, \mathbf{p}) + \frac{1}{2} \sum_j^d \frac{\partial^2 \Phi}{\partial x_j \partial x_k} \frac{\partial^2 J}{\partial p_j \partial p_k}(s, \mathbf{x}, \mathbf{p}) \right]. \quad (45)$$

Note that (45) requires the Hessian matrix of  $\Phi$ ,

$$Z_{j,k} \equiv \frac{\partial^2 \Phi}{\partial x_j \partial x_k}. \quad (46)$$

On characteristics,  $Z$  satisfies the Ricatti equation [18, 20],

$$\frac{dZ}{dt} = -ZDZ - ZC - C^T Z - G, \quad (47)$$

where

$$D_{j,k}(\mathbf{x}, \mathbf{p}) = \frac{\partial^2 \mathcal{H}}{\partial p_j \partial p_k}, \quad C_{j,k}(\mathbf{x}, \mathbf{p}) = \frac{\partial^2 \mathcal{H}}{\partial p_j \partial x_k}, \quad G_{j,k}(\mathbf{x}, \mathbf{p}) = \frac{\partial^2 \mathcal{H}}{\partial x_j \partial x_k}. \quad (48)$$

## C Numerical algorithm to compute the principal eigenvalue

Power iteration is perhaps the simplest eigenvalue algorithm, and if it converges, it converges to the eigenvalue with the largest absolute value. However, we seek the eigenvalue with the largest real part, which is not necessarily the eigenvalue with the largest absolute value. Let  $\kappa > 0$  be a shift of the spectrum of  $Q$ , and define the matrix

$$M_\kappa \equiv Q + \kappa I. \quad (49)$$

For  $\kappa$  sufficiently large, the matrix  $M_\kappa$  is nonnegative and irreducible. Both of these properties follow from the structure of the transition rate matrix  $A^x$ , which we assume to be irreducible. It follows from the Perron–Frobenius Theorem that  $M_\kappa$  has a simple real eigenvalue, call it  $\mu$ , that is greater in absolute value than the remaining eigenvalues. The eigenvector corresponding to  $\mu$ , call it  $r$ , is strictly positive. Setting  $\mu = \kappa + \lambda$ , we have that  $M_\kappa r = Qr + \kappa r = \mu r = \kappa r + \lambda r$ . It follows that  $\lambda$  and  $r$  are an eigenpair of  $Q$ . Moreover,  $\lambda$  is the principal eigenvalue we seek. We can then apply power iteration to  $M_\kappa$ , and the result converges to  $\mu$  from which we obtain our desired result with  $\lambda = \mu - \kappa$ . It remains to determine the convergence rate of the scheme and what value to choose for  $\kappa$ .

There are two relevant properties of nonnegative irreducible matrices. Let  $z$  be an arbitrary positive vector. First, we have that  $(M_\kappa z)_i > 0$ . That is, the matrix maps positive vectors to positive vectors. Second, it follows from Theorem 8.1.26 in Ref. [13] that the eigenvalue  $\mu$  is bounded by

$$\min_i \frac{(M_\kappa z)_i}{z_i} \leq \mu \leq \max_i \frac{(M_\kappa z)_i}{z_i}. \quad (50)$$



Moreover, if we optimize the bound over all possible positive vectors, we have

$$\max_{z>0} \min_i \frac{(M_\kappa z)_i}{z_i} = \min_{z>0} \max_i \frac{(M_\kappa z)_i}{z_i} = \frac{(M_\kappa r)_i}{r_i} = \frac{\mu r_i}{r_i} = \mu. \quad (51)$$

Let  $\hat{\mu}$  be the next largest eigenvalue of  $M$  in absolute value, and let  $\hat{\mu} = \kappa + \hat{\lambda}$ . The asymptotic convergence rate is [30]

$$|\lambda - \lambda_n| = |\mu - \mu_n| \sim O\left(\left|\frac{\mu}{\hat{\mu}}\right|^{2n}\right) = O\left(\left|\frac{\lambda + \kappa}{\hat{\lambda} + \kappa}\right|^{2n}\right), \quad n \gg 1.$$

It follows that as  $\kappa \rightarrow \infty$ , the convergence rate approaches unity. We therefore want to choose the smallest possible value for  $\kappa$ , which corresponds to  $\kappa = -\min_i Q_{ii}$ . Notice that the shift could be negative if all of the diagonal entries of  $Q$  are positive, which corresponds to the case where  $Q$  is already a nonnegative matrix. No shift is therefore necessary, but the negative shift speeds up convergence since  $M_\kappa$  is also nonnegative. The algorithm written in Python code (using Scipy) as follows.

```
from pylab import *
def collatz(Q, err_tol):
    ## Input: square matrix Q with dimension N
    ##          error tolerance, err_tol
    ## Output: principal eigenvalue approximation, mu
    N = M.shape[0]
    M = Q + kappa*eye(N)
    z = ones((N, 1))
    error = err_tol + 1.
    while (error > err_tol):
        q = dot(M, z)
        d = q/z
        a = min(d)
        b = max(d)
        mu = 0.5*(a + b)
        z = q/mu
        error = b - a
    return mu - kappa
```

## References

- [1] M. ASSAF, E. ROBERTS, AND Z. LUTHEY-SCHULTEN, *Determining the stability of genetic switches: Explicitly accounting for mrna noise*, Phys. Rev. Lett., 106 (2011), p. 248102.
- [2] E. AURELL AND K. SNEPPEN, *Epigenetics as a first exit problem*, Phys. Rev. Lett., 88 (2002), p. 048101.
- [3] P. C. BRESSLOFF AND J. M. NEWBY, *Path integrals and large deviations in stochastic hybrid systems*, Phys. Rev. E, 89 (2014), p. 042701.
- [4] M. CAMERON, *Finding the quasipotential for nongradient sdes*, Physica D, 241 (2012), pp. 1532 – 1550.
- [5] C. R. DOERING, K. V. SARGSYAN, L. M. SANDER, AND E. VANDEN-EIJNDEN, *Asymptotics of rare events in birth–death processes bypassing the exact solutions*, Journal of Physics: Condensed Matter, 19 (2007), p. 065145.
- [6] M. DOI, *Stochastic theory of diffusion-controlled reaction*, J. Phys. A: Math. Gen., 9 (1976), p. 1479.

- [7] A. ELДАР AND M. B. ELOWITZ, *Functional roles for noise in genetic circuits*, Nature, 467 (2010), pp. 167–173.
- [8] H. D. FENG, K. ZHANG, AND J. WANG, *Non-equilibrium transition atate rate theory*, J. Chem. Sci., (2014).
- [9] J. FENG AND T. G. KURTZ, *Large deviations for stochastic processes*, vol. v. 131 of Mathematical surveys and monographs, American Mathematical Society, Providence, R.I., 2006.
- [10] M. I. FREIDLIN AND A. D. WENTZELL, *Random Perturbations of Dynamical Systems*, Springer- Verlag, Berlin Heidelberg, 3rd edition ed., 2012.
- [11] C. W. GARDINER, *Handbook of stochastic methods for physics, chemistry, and the natural sciences*, vol. v. 13, Springer-Verlag, Berlin, 1983.
- [12] M. HEYMANN AND E. VANDEN-ELJNDEN, *The geometric minimum action method: A least action principle on the space of curves*, Communications on Pure and Applied Mathematics, 61 (2008), pp. 1052–1117.
- [13] R. A. HORN AND C. R. JOHNSON, *Matrix analysis*, Cambridge University Press, Cambridge, 1985.
- [14] J. P. KEENER AND J. M. NEWBY, *Perturbation analysis of spontaneous action potential initiation by stochastic ion channels*, Phys. Rev. E, 84 (2011), p. 011918.
- [15] T. B. KEPLER AND T. C. ELSTON, *Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations*, Biophys. J., 81 (2001), pp. 3116–3136.
- [16] Y. KIFER, *Large deviations and adiabatic transitions for dynamical systems and markov processes in fully coupled averaging*, Memoirs of the American Mathematical Society, 201 (2009), pp. 1–144.
- [17] M. LU, J. ONUCHIC, AND E. BEN-JACOB, *Construction of an effective landscape for multistate genetic switches*, Phys. Rev. Lett., 113 (2014), p. 078102.
- [18] D. LUDWIG, *Persistence of dynamical systems under random perturbations*, SIAM Review, 17 (1975), pp. pp. 605–640.
- [19] N. MAHESHRI AND E. K. O’SHEA, *Living with noisy genes: How cells function reliably with inherent variability in gene expression*, Annu. Rev. Biophys. Biomol. Struct., 36 (2007), pp. 413–434.
- [20] R. S. MAIER AND D. L. STEIN, *Limiting exit location distributions in the stochastic exit problem*, SIAM J. Appl. Math., 57 (1997), pp. 752–790.
- [21] J. NEWBY, *Bistable switching asymptotics for the self regulating gene*, ArXiv e-prints, 1407.4344 (2014), pp. 1–9.
- [22] ———, *Spontaneous excitability in the morris–lecar model with ion channel noise*, ArXiv e-prints, 1406.2914 (2014), pp. 1–47.
- [23] J. NEWBY AND J. CHAPMAN, *Metastable behavior in markov processes with internal states*, J. Math. Biol., (2013), pp. 1–36.
- [24] J. M. NEWBY, *Isolating intrinsic noise sources in a stochastic genetic switch*, Physical Biology, 9 (2012), p. 026002.
- [25] J. M. NEWBY AND J. P. KEENER, *An asymptotic analysis of the spatially inhomogeneous velocity-jump process*, Multiscale Model. Simul., 9 (2011), pp. 735–765.

- [26] J. OCKENDON, S. HOWISON, A. LACEY, AND A. MOVCHAN, *Applied partial differential equations*, Oxford University Press, Oxford, rev. ed ed., 2003.
- [27] L. PELITI, *Path integral approach to birth-death processes on a lattice*, J. Phys. France, 46 (1985), pp. 1469–1483.
- [28] M. SASAI AND P. G. WOLYNES, *Stochastic gene expression as a many-body problem*, Proc. Natl. Acad. Sci. U.S.A., 100 (2003), pp. 2374–2379.
- [29] Z. SCHUSS, *Theory and applications of stochastic processes: an analytical approach*, vol. v. 170 of Applied mathematical sciences, Springer, New York, 2010.
- [30] L. N. TREFETHEN, *Pseudospectra of linear operators*, SIAM Review, 39 (1997), pp. 383–406.
- [31] E. VANDEN-ELJNDEN AND M. HEYMANN, *The geometric minimum action method for computing minimum energy paths*, The Journal of Chemical Physics, 128 (2008).
- [32] A. M. WALCZAK, J. N. ONUCHIC, AND P. G. WOLYNES, *Absolute rate theories of epigenetic stability*, Proc. Natl. Acad. Sci. U.S.A., 102 (2005), pp. 18926–18931.
- [33] B. ZHANG AND P. G. WOLYNES, *Stem cell differentiation as a many-body problem*, Proc. Natl. Acad. Sci. U.S.A., 111 (2014), pp. 10185–10190.